

ВІДГУК

офіційного опонента Терейковського Ігоря Анатолійовича
на дисертаційну роботу Долгих Сергія Миколайовича
«Інформаційна технологія розпізнавання мережевих даних Інтернет на основі
генеративних нейромережевих моделей»,
що подана на здобуття наукового ступеня кандидата технічних наук
за спеціальністю 05.13.06 – Інформаційні технології

Актуальність теми дисертаційної роботи

В сучасних умовах увага науковців багатьох країн світу прикута до проблематики аналізу та класифікації даних Інтернет, що має важливе значення як для забезпечення ефективного функціонування різноманітних інформаційних систем, так і для забезпечення достатнього рівня їх захисту. Відсутність явних алгоритмічних рішень, складність піддослідних процесів, наявність суб'єктивних факторів обумовили інтерес до застосування рішень, котрі базуються на методах машинного інтелекту. При цьому серйозною перешкодою до розробки відповідних інформаційних технологій являється обмеженість доступних репрезентативних наборів навчальних даних для розробки інформаційних систем розпізнавання трафіку Інтернет. Крім того, в багатьох випадках навіть за наявності навчальних даних успіх навчання не може бути гарантований, тому що параметри трафіку в різних мережах можуть мати значну варіацію і успішне навчання на даних, отриманих з «базової» мережі, може не забезпечити успіху з реальними даними з іншої мережі. Успішне застосування методів навчання систем машинного інтелекту у таких задачах та галузях повинні враховувати особливості розподілу параметрів у конкретних локальних мережах, де формування достатніх маркованих навчальних наборів в багатьох випадках є занадто довготривалим та вимагає значних витрат. В результаті виникає задача розробки інформаційних систем та технологій класифікації даних Інтернет, що базуються на засобах штучного інтелекту, здатних навчатися на немаркованих даних, не використовуючи системи заздалегідь відомих концептуальних структур; використовувати локальні, натуральні особливості та структури даних для успішного навчання; забезпечуючи стабільний успіх при навчанні з обмеженими наборами даних. Розробка таких підходів потребує теоретичного обґрунтування, експериментальної перевірки та

оцінки ефективності нових методів та моделей навчання з мінімальною кількістю навчальних даних. Відповідно до сучасних досягнень в області штучного інтелекту, в якості базису створення означених підходів, моделей та методів доцільно використовувати апарат генеративних нейромережевих моделей. Тому тема дисертаційної роботи, що пов'язана з побудовою інформаційної технології розпізнавання мережевих даних Інтернет на основі генеративних нейромережевих моделей, має безумовну актуальність і значимість.

Обґрунтованість наукових положень, висновків і рекомендацій, сформульованих у дисертації

Ступінь обґрунтованості є достатньою і забезпечена коректним застосуванням методів дослідження, що ґрунтуються на методологічному базисі теорії інформації, статистичних методах, теорії неконтрольованого навчання та представлень – для розробки методів створення інформативних представлень даних Інтернет зі значно зниженою розмірністю; моделях та методах неконтрольованого навчання, в тому числі генеративних моделях нейронних мереж автоенкодера – для створення інформативних представлень даних зі значно зниженою розмірністю; методах кластеризації, включаючи кластеризацію за щільністю – для аналізу структури неконтрольованих генеративних представлень; методах статистичного аналізу – для обробки результатів експериментів та верифікації ефективності розроблених методів.

Зв'язок дисертаційної роботи з пріоритетними напрямками розвитку науки і техніки, державними чи галузевими науковими програмами

Результати науково-дослідної роботи використано в Національному авіаційному університеті при виконанні держбюджетних науково-дослідних робіт за темами:

– №247-ДБ19 «Розроблення та виготовлення програмно-апаратних засобів цільового навантаження для повітряного спостереження та альтернативної навігації літального апарату» (№ держреєстрації: 0119U100553);

– № 421-ДБ22 «Інтелектуалізована система захищеного передавання пакетних даних на базі розвідувально-пошукового безпілотного літального апарату» (№ держреєстрації: 0122U002361).

Достовірність одержаних результатів, повнота їх викладу в опублікованих працях

В дисертаційній роботі достатньо чітко та лаконічно визначаються мета, об'єкт, предмет дослідження та формулюються завдання для досягнення поставлених цілей. Достовірність наукових положень, які наведені в дисертаційній роботі, обумовлені доволі грамотною постановкою та комплексним підходом до вирішення завдань дослідження. Методологічно обґрунтовані вихідні положення роботи, які підтверджуються практичними результатами і забезпечується коректним опрацюванням отриманих експериментальних даних, їх кількісним і якісним аналізом, що надає отриманим науковим результатам роботи необхідної переконливості і практичної цінності. Отримані в дисертації результати впродовж 2019-2022 рр. оприлюднено та обговорено на 14 наукових конференціях. Результати дисертаційної роботи з необхідною повнотою викладено у 21 науковій праці, в тому числі 1 розділі колективної монографії та 17 наукових статях (16 – у міжнародних виданнях, що входять до БД Scopus, 1 – у вітчизняних фахових наукових журналах категорії Б). Детальне ознайомлення зі змістом публікацій Долгих С.М. свідчить про повноту викладення результатів дисертації у міжнародних та вітчизняних наукових фахових виданнях. Автореферат відповідає змісту та основним положенням дисертації.

Найбільш суттєві наукові результати дисертації

1. Удосконалено концептуальну і математичну модель розподілів даних пакетів трафіку Інтернет, що дозволяє підвищити ефективність виділення інформативних факторів даних. Новизна моделей полягає у застосуванні неконтрольованих генеративних представлень даних пакетів трафіку Інтернет.

2. Отримали подальший розвиток методи теорії генеративних представлень, що забезпечили підвищення ефективності навчання за характеристиками зниження помилки та відтворення даних при неконтрольованому генеративному навчанні з використанням даних Інтернет та інших типів даних. Новизна методів полягає у застосуванні оригінальної архітектури автоенкодера з різким стисненням розмірності шару кодування.

3. Вперше доведена теорема про категоризацію генеративних представлень, що на підставі методів варіаційного аналізу забезпечує

теоретичне обґрунтування методів навчання на основі генеративної структури представлень.

4. Вперше розроблено методи навчання призначені для розпізнавання відомих класів даних пакетів трафіку Інтернет, які забезпечують: достатню точність розпізнавання; зменшення залежності від джерела отримання навчальних даних; зменшення обсягу навчальних даних в порівнянні з відомими методами контрольованого навчання. Новизна методів обумовлена використанням запропонованих концептуальної та математичної моделі, методів теорії генеративних представлень та теореми про категоризацію генеративних представлень.

5. Вперше розроблено методи навчання призначені для розпізнавання натуральних концептів даних пакетів трафіку Інтернет, що на відміну від відомих за рахунок визначення структури щільності генеративних представлень, дозволяють реалізовувати розпізнавання натуральних концептів даних Інтернет без використання маркованих даних, з точністю на рівні відомих методів контрольованого навчання.

Наукова новизна дисертаційної роботи обумовлена рішенням завдань розробки ефективних методів машинного інтелекту здатних навчатися з мінімальними наборами даних Інтернет. Результати роботи є науково обґрунтовані і не суперечать існуючим уявленням. Новизна наукових положень, висновків та рекомендацій, сформульованих у дисертації, підтверджується відсутністю у літературі аналогічних підходів.

Практичне значення одержаних в дисертаційному дослідженні результатів

1. Визначено значення показників ефективності сучасних методів розпізнавання та класифікації даних пакетів трафіку Інтернет.

2. Розроблено програмну реалізацію моделей генеративного навчання автоенкодера з різким стиском розмірності шару кодування, що дозволяє отримати інформативні представлення даних Інтернет суттєво зниженої розмірності.

3. Розроблено програмне втілення масивів даних пакетів трафіку Інтернет та зображень.

4. Розроблено програмну реалізацію методів визначення структури щільності генеративних представлень, що забезпечує стабільне визначення

структури щільності генеративних представлень даних Інтернет.

5. Розроблено програмну реалізацію методу навчання розпізнавання відомих класів даних пакетів трафіку Інтернет з використанням структури щільності генеративних представлень, що забезпечує стабільне навчання розпізнавання відомих класів даних Інтернет при зменшенні обсягу навчальних даних, у 10–100 разів в порівнянні з відомими методами контрольованого навчання.

6. Розроблено програмну реалізацію методу навчання розпізнавання натуральних концептів даних пакетів трафіку Інтернет з використанням структури щільності генеративних представлень, що забезпечує стабільне навчання розпізнавання натуральних типів даних Інтернет без використання маркованих даних.

7. Отримано програмне прототипне виконання інформаційної технології розпізнавання даних пакетів трафіку Інтернет при наборах навчання мінімального обсягу з використанням структури щільності генеративних представлень.

Оцінка змісту дисертації, її завершеність в цілому

Дисертація містить анотацію українською та англійською мовами, список публікацій здобувача (21 публ.), в яких опубліковані основні наукові результати дисертації. З них 1 розділ колективної монографії, 17 наукових статей (16 – у міжнародних рецензованих виданнях, що входять до бази даних Scopus, 1 – у вітчизняних фахових наукових журналах категорії Б), а також 3 матеріали тез доповідей на конференціях.

Основна частина дисертаційної роботи складається зі вступу, чотирьох розділів, що містять основні наукові результати, загальних висновків, списку використаної літератури та двох додатків.

У вступі розкрито сутність і стан наукової задачі та її значущість, обґрунтовано необхідність проведення дисертаційних досліджень. В ньому подана загальна характеристика роботи у відповідності з діючими вимогами до дисертацій.

Перший розділ присвячено аналізу сучасних методів та моделей неконтрольованого навчання та розподілів даних у багатовимірних просторах представлень. На основі огляду літератури за обраною тематикою проведено визначення основних понять, що використовуються у дисертаційному

дослідженні. Визначено характеристики сучасних методів, підходів та моделей неконтрольованого. Розглянуто методи та критерії визначення інформаційної відповідності розподілів випадкових величин. Проаналізовано методи, програмні продукти, бібліотеки обробки даних та машинного навчання, а також експериментальні результати у неконтрольованому навчанні. У результаті проведеного аналізу обґрунтовано, що використання генеративних нейромережових архітектур може бути ефективним підходом у створенні інформативних представлень даних та являється перспективним шляхом розв'язання поставленого завдання успішного навчання штучних систем інтелекту з мінімальними відомими даними. Також визначено, що вирішення задач дисертаційного дослідження потребує вдосконалення теоретичних основ систем генеративного навчання.

У *другому розділі* вирішувалась наукова задача подальшого розвитку теоретичних основ моделей та методів у галузі неконтрольованого навчання неконтрольованих представлень моделей самокодування та генерації, даних та методів навчання моделей, що використовувалися або мають істотне відношення до напрямку дослідження, методів та результатів роботи. У розділі наведено визначення основних концепцій теорії неконтрольованого навчання, розглянуто принципи та методи побудови та навчання моделей неконтрольованого навчання, створення та аналізу неконтрольованих представлень та доведено ключові результати для обґрунтування принципів та методів побудови моделей неконтрольованого навчання. На підставі отриманих теоретичних результатів обґрунтовано та розроблено методи навчання систем на основі структури щільності (ландшафту) генеративних представлень даних пакетів трафіку Інтернет. Також проведено аналіз отриманих результатів, зроблено висновки, що пов'язують результати роботи з іншими областями теорії систем машинного навчання та неконтрольованого навчання, а також запропоновано гіпотези та зроблено припущення для експериментальної перевірки.

У *третьому розділі* наведено детальний опис генеративних архітектур, моделей, методів та даних, використаних в роботі. На підставі результатів теоретичного обґрунтування, а також опублікованих наукових результатів, була обрана архітектура моделей, що використовуються для створення та вивчення генеративних неконтрольованих представлень вхідних даних. Особливої уваги заслуговує метод розпізнавання відомих класів на основі

генеративного ландшафту з мінімальними наборами навчання та метод неконтрольованого розпізнавання натуральних концептів даних на основі генеративного ландшафту.

Четвертий розділ присвячено експериментальній перевірці запропонованих у роботі теоретичних висновків та методів навчання з використанням генеративної латентної структури (ландшафту) даних Інтернет. Також в розділі обґрунтовано побудову інформаційної технології навчання з мінімальними даними навчання з використанням генеративної структури (ландшафту) представлень, що є логічним узагальненням запропонованих моделей та методів. Оригінальною складовою розробленої технології є узагальнення процесів навчання на основі генеративної структури представлень та забезпечення вимоги щодо мінімізації обсягів наборів відомих даних навчання до окремих зразків відомих класів. В розділі наведено приклади експериментальних досліджень, які показують ефективність запропонованих методів та інформаційної технології, а відповідно, і підтверджують достовірність основних результатів дисертації.

У висновках викладено найважливіші наукові та практичні результати, отримані в дисертації.

Список використаних джерел оформлений в основному правильно, складається з 110 найменувань наукової літератури за темою дисертації.

Додатки містять діаграму архітектури моделей глибокого автоенкодера із стисненням та акт впровадження результатів дисертаційної роботи.

Загальний обсяг дисертації становить 139 сторінок, матеріал дисертації викладений на високому науковому рівні, а стиль викладення матеріалу забезпечує доступність його сприйняття і відповідає існуючим вимогам до наукових праць. Тема дисертації та її зміст відповідають напрямку досліджень паспорту спеціальності 05.13.06 – Інформаційні технології.

Отримані в дисертації нові теоретичні положення доцільно використовувати в наукових дослідженнях та практичних розробках науково-дослідних організацій, робота яких пов'язана не тільки з аналізом мережевого трафіку, але й в області де існує необхідність використання систем машинного інтелекту з мінімальними даними істини (захист інформації, медицина).

Зауваження щодо дисертації

1. Деяка невідповідність використаної в дисертації термінології класичним означенням дещо ускладнює сприйняття результатів, отриманих автором.

2. Автору варто було б більш докладно описати механізми визначення структурних параметрів нейромережових моделей, використаних в роботі.

3. При розгляді нейронних мереж, перспективних з точки зору використання в системах неконтрольованого навчання, недостатня увага приділена карті Кохонена.

4. В тексті дисертації не досить чітко визначені обмеження на використання запропонованих автором методів навчання з використанням генеративної структури (ландшафту) представлення.

5. Поза межами дисертаційного дослідження залишились задачі, пов'язані з формуванням баз даних параметрів трафіку Інтернет.

6. До оформлення тексту дисертації є окремі технічні зауваження (на сторінці 20 в одних дужках забагато посилань на літературні джерела, сторінка 80 містить всього 4 рядки тексту).

Проте вказані недоліки не знижують наукової і практичної цінності нових наукових результатів дисертаційних досліджень, і не впливають на її загальну позитивну оцінку.

Оцінка відповідності дисертації встановленим вимогам

Дисертація є завершеною кваліфікаційною роботою, містить отримані автором нові науково обґрунтовані результати, які у сукупності забезпечили створення ефективних методів забезпечення стабільності та ефективності навчання систем машинного інтелекту з мінімальними даними істини з використанням природної структури даних Інтернет та розробку на цій основі інформаційної технології розпізнавання мережових даних Інтернет.

Результати дисертаційної роботи у галузі розробки інформаційних технологій аналізу мережевого трафіку є пріоритетними.

ВИСНОВОК

Вважаю, що дисертаційна робота Долгих С. М. є завершеним науковим дослідженням, направленим на вирішення актуальної науково-прикладної задачі розробка і дослідження нових ефективних методів забезпечення

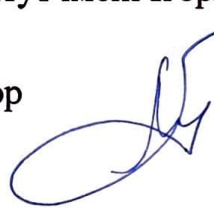
стабільності та ефективності навчання систем машинного інтелекту з мінімальними даними істини з використанням природної структури даних Інтернет. Дисертація має теоретичне та практичне значення, виконана із дотриманням принципів академічної доброчесності, відповідає паспорту спеціальності 05.13.06 – Інформаційні технології та вимогам «Порядку присудження наукових ступенів», а її автор Долгих Сергій Миколайович заслуговує на присудження наукового ступеня кандидата технічних наук за спеціальністю 05.13.06 – Інформаційні технології.

Офіційний опонент

професор кафедри системного програмування
і спеціалізованих комп'ютерних систем
Національного технічного університету України
"Київський політехнічний інститут імені Ігоря Сікорського"

доктор технічних наук, професор

5 травня 2023 р.



Ігор ТЕРЕЙКОВСЬКИЙ

Особистий підпис Терейковського Ігоря Анатолійовича засвідчую:

вчений секретар

Національного технічного університету України
"Київський політехнічний інститут імені Ігоря Сікорського"

кандидат технічних наук, доцент

03 травня 2023 р.



Валерія ХОЛЯВКО